

## How decisions about fitting species distribution models affect conservation outcomes

Angela Muscatello<sup>1\*</sup>, Jane Elith<sup>1</sup> and Heini Kujala<sup>1,2</sup>

<sup>1</sup> School of BioSciences, The University of Melbourne, Parkville VIC 3010, Australia

<sup>2</sup> Finnish Museum of Natural History, University of Helsinki, FI-00140 Helsinki, Finland

\* Corresponding author

Email: [angelamuscatello1@gmail.com](mailto:angelamuscatello1@gmail.com) (AM)

School of BioSciences, The University of Melbourne, Parkville VIC 3010, Australia

Running head: SDMs in conservation planning

Word count: 6052

Keywords

Species distribution models, spatial prioritisation, conservation planning, land use planning, uncertainty, observation bias, thresholding

Article Impact Statement

Methodological decisions on how species distribution models (SDMs) are built can alter locations of priority areas up to 90%.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/cobi.13669](https://doi.org/10.1111/cobi.13669).

This article is protected by copyright. All rights reserved.

## Abstract

Species distribution models (SDMs) are increasingly used in conservation and land use planning as inputs to describe biodiversity patterns. SDMs can be built in different ways, and decisions about data preparation, selection of predictor variables, model fitting and evaluation all alter the resulting predictions. Commonly, the true distribution of species is not known, nor is there independent data to verify which SDM variant to choose. Such model uncertainty is concerning to planners. We analysed how 11 routine decisions about model complexity, predictors, bias treatment and thresholding of predicted values altered conservation priority patterns across 25 species. While all SDM variants had good model performance ( $AUC > 0.7$ ), they produced spatially different predictions for species and different conservation priority solutions. Priorities were most strongly altered by decisions to not deal with bias or to apply binary thresholds to predicted values, where on average 40% and 35%, respectively, of all grid cells received an opposite priority ranking. Forcing high model complexity altered conservation solutions less than forcing simplicity (14% and 24% of cells with opposite rank values, respectively), while using fewer species records to build models or choosing alternative bias treatments had intermediate effects (25% and 23%, respectively). Depending on modelling choices, priority areas overlapped as little as 10-20% with the baseline solution, affecting top and bottom priorities differently. Our results demonstrate the extent of model-based uncertainty and quantify the relative impacts of SDM building decisions. When the truth about the best SDM approach and conservation plan is not known, solving uncertainty or spending time considering alternative options is most important for those decisions that change plans the most.

## Introduction

Conservation of biodiversity requires decisions on how to allocate resources in space, in order to protect, restore or manage areas within a landscape. This can be challenging, typically requiring

spatial data and considerations across many candidate conservation sites and tens to hundreds of species. Spatial optimisation tools such as Marxan (Ball, Possingham & Watts 2009) and Zonation (Moilanen et al. 2005) help planners and managers understand how resources can be cost-effectively distributed between candidate locations so as to maximise conservation benefits for multiple species (Kujala et al. 2018a).

Species distribution models (SDMs) expand the spatial coverage of species information from point observations to mapped predictions across landscapes and are thus useful for conservation planning (Guisan et al. 2013). Building SDMs requires decisions at each step of data preparation, selection of predictor variables and modelling method, model fitting, and evaluation. Different decisions alter predictions, sometimes substantially, exposing uncertainty in predicted distribution patterns. Previous research has explored several sources of SDM error and uncertainty (e.g. Barry et al. 2006; Langford et al. 2009; Synes & Osborne 2011; Hermoso & Kennard 2012). Yet, there is still considerable variation in how SDMs are built, partly because no single approach is proven best across applications (Araujo et al. 2019).

In the context of conservation planning, errors in SDM predictions can result in excessive spending of resources or the neglect of species in areas mistaken as unimportant (Wilson et al. 2006; Guisan et al. 2013). As species' true distributions are not known, we cannot fully evaluate errors in alternative SDM predictions. Modellers can tune models against held-out data, but (particularly when data are biased) this may not get closer to truth (El-Gabbas and Dormann 2018). Instead, when no additional information motivates choice between SDM variants, these become alternative representations of species' true distributions, their differences illustrating the uncertainty in our understanding of where species occur. Despite their use by decision makers (Guisan et al. 2013) and their likely impact on conservation outcomes (Wilson et al. 2006; Guillerá-Arroita et al. 2015), uncertainty in SDMs is rarely reported or systematically assessed (Langford et al. 2009; Tulloch et al. 2016). There also

exists little guidance on how relevant these uncertainties are for conservation decisions, particularly multi-species ones, and whether some uncertainties are more important to resolve than others.

Here we present a systematic evaluation of how decisions made during SDM building alter conservation plans. We do not focus on creating or identifying the best possible model, but rather on representing common choices of conservation practitioners. We used real data from eastern Australia, to quantify the relative influence these methodological decisions have on i) the spatial patterns in species' predicted distribution; ii) the spatial patterns of conservation priorities; and iii) the overlap of most and least important areas for conservation. Understanding how stable conservation plans are against different SDM alterations will help identify those uncertainties that are most critical to solve, and which decisions may need to be more carefully considered.

## Methods

### Study area

The Greater Hunter region covers 29,145 km<sup>2</sup> along the eastern coast of Australia, within New South Wales (NSW). It's habitats (Fig. 1), range from coastal wetlands and inland grasslands to sclerophyll forests and highland rainforest. Large parts of the region support agriculture and mining. Altitudes vary from sea level to 1,596 m at Brumlow Top (Fig.1).

### Species data

We used location records for 25 native species listed as protected under NSW law, covering four broad taxonomic groups: 11 plant, 7 bird, 4 mammal and 3 reptile species (Supporting Information). These were a subset of species data cleaned and collated by Kujala et al. (2015), originally sourced from the Atlas of Living Australia (<https://www.ala.org.au/>) and BioNet (<http://www.bionet.nsw.gov.au/>). Species were selected for their influence in defining important biodiversity locations in multi-species conservation prioritisations for the Greater Hunter (Kujala et al. 2018b). Species records were collected after 1 January 1990 and had recorded accuracy of 100

This article is protected by copyright. All rights reserved.

meters or less. Duplicate records were removed. For details of data cleaning and processing, see Kujala et al. (2015).

### **Environmental data**

We gathered a set of 24 environmental variables, describing aspects of the climate, soil, topography and vegetation at 100 m resolution, as candidate variables for the SDMs (Supporting Information, including tests of pairwise correlations). Kujala et al. (2015) identified these variables as ecologically relevant for explaining the observations of species.

### **Species Distribution Modelling**

We used MaxEnt (Phillips, Anderson & Schapire 2006; Phillips & Dudik 2008) version 3.4.1 to create several species distribution models (SDMs) for each species. The models were fitted with the *dismo* package (version 1.1-4) in R (version 3.5.0), using sites with data mode. Predictions were made at 100 m resolution over the whole region. We first created a baseline SDM for each species (hereafter *base*), fitting the model to all cleaned observation data. Since we do not know truth, we do not assert these settings provide a distribution closest to the truth, rather it provides a reasonably motivated point of comparison. We aimed for settings that are arguably sound practice, commonly used and based on good reasoning (Table 1 and Supporting Information). These included MaxEnt's defaults for model complexity, in which programmed rules enforce varying complexity of modelled responses for different numbers of species records. These defaults, which produce moderately complex models, are justifiable when predicting to the same region as the training data (Merow et al. 2014). Biases in the species data were addressed with target group background (TGB) samples, which quantify survey effort through observations of multiple species in the same taxonomic group (Phillips et al. 2009; Supporting Information). Following Kujala et al. 2015 and motivated by the desire to use ecologically relevant predictors, we used different subsets of the 24 candidate predictors for each of the four taxonomic groups (Supporting Information).

We explored 11 variations of the base model for producing distribution maps for species. Our first variations focused on model complexity within MaxEnt. These mimic likely differences across modelling methods, including different SDM algorithms. MaxEnt controls model complexity through feature types and regularisation parameters (Elith et al. 2011; Merow et al. 2013); here we focus on features. In the *simple* variant, feature types for all species were restricted to linear and quadratic only. Whilst not a common choice for many MaxEnt users, we use it here to mimic a regression model with linear and quadratic terms. In the *complex* variant, use of all available feature types was enforced for all species. This includes use of threshold features (omitted from default settings) and mimics a complex machine learning approach where fitted functions cannot be simplified (Elith 2019).

Second, we focussed on predictor variable selection. The base used the 16-19 most ecologically relevant predictors for each taxonomic group (Supporting Information). For the *predictors* variant, the same constant set of 12 variables were used for all species. These included the first two axes of a principal components reduction (via *RStoolbox* version 0.2.1) of the 6 climate variables, plus most remaining variables but excluding some vegetation variants (Supporting Information).

Next, we focused on alternative methods for handling expected biases in species data (Phillips et al. 2009). Three *no bias* models were produced, replacing the TGB with 2000, 10000 or 30000 randomly allocated background points. These mimic the common choice to ignore bias and explore uncertainties around how many samples to use as background points (Renner et al. 2015). A fourth approach (*bias grid*) dealt with bias, but through bias grids (Kujala et al. 2015) instead of TGB points, for each major taxonomic group. Each bias grid was created with a normalised density kernel fitted to all taxon-specific target group records (Supporting Information) with a 10 km radius. The grid values provided probability weights for sampling 20000 background points, meaning that relatively more background points were sampled from regions with higher densities of records.

We also explored the impact of having fewer species records. In the *few records* variant, the number of presence points was reduced by half for each species. Since results may depend on the random sample selected, we ran 10 repeats per species.

The *typical* model variant reflected what we often see in conservation planning papers (e.g. Kaky & Gilbert 2016), with a constant set of predictor variables across species, no bias treatment, and 10000 random background samples.

Our final alterations explore the common approach of thresholding predicted values. We used the equate entropy threshold in MaxEnt, which chooses the threshold giving the thresholded distribution the same entropy as the original output. This threshold was used to make the *binary* variant, assigning a value of 0 below the threshold and 1, above. A second *truncated* variant kept the continuous predictions above the threshold but assigned 0 below it. This aims to eliminate cells likely unsuitable for the species, whilst retaining predictive detail for more suitable sites.

In total, we produced 19 SDMs for each species: 1 base model, 8 variations to model building, plus 10 iterations with 50% species records. The two threshold variants were applied to each SDM (Table 1). All modelling code and data can be accessed at: <https://figshare.com/s/73725332d2ed92cbf9c1>

### **Measuring SDM differences**

As a point-based measure of model performance we recorded the area under the receiver operating characteristic curve (AUC) estimated for presence vs background samples (Foody 2011), a measure of model discrimination. We ran a 7-fold cross validation for each species' SDM across all model variants, estimating  $AUC_{train}$  across the training folds, and  $AUC_{test}$  across the test folds.

We also quantitatively compared the mapped outputs. For each species, we compared all pairwise combinations of model variants using a Pearson's pairwise correlation coefficient and an asymmetric Kulczynski's coefficient (KUL), which was calculated as

**Equation 1.**

$$KUL_{xy} = \frac{\sum_{i=1}^n x_i - \sum_{i=1}^n \min(x_i, y_i)}{\sum_{i=1}^n x_i}$$

(Legendre & Legendre 1998; Ray & Burgman 2006). This measure focusses on agreement of high values, asking whether the models identify best habitat in the same locations. Values close to 0 indicate similar patterns. KUL can be calculated from the viewpoint of model  $x$  or  $y$ , and  $KUL_{xy}$  is not, in general, equal to  $KUL_{yx}$ . For instance, while two model variants may predict highest values in the same cells,  $x$  may assign zero to most other cells, while  $y$  may have a greater proportion of intermediate ranked cells. In this case,  $KUL_{xy}$  will be lower than  $KUL_{yx}$ .

### Spatial prioritisations

To explore how SDM decisions propagate into conservation plans, we produced several conservation prioritisations, each based on one SDM variant (Table 1). We used the spatial prioritisation software Zonation (version 4.0.0) to rank the grid cells according to their biodiversity value (Moilanen et al. 2005). The program starts by assuming the entire landscape is protected, then iteratively removes grid cells with the smallest marginal ( $\delta$ ) value for included species, until no cells remain. The removal order defines the priority rank of each cell, the most important grid cells being removed last. We ran Zonation with the core area option (Moilanen et al. 2014) which defines the marginal loss value  $\delta$  of cell  $i$  at each iterative step as

**Equation 2.**

$$\delta_i = \max_j \frac{p_{ij}}{\sum_{k \in S} p_{kj}}$$



where  $p_{ij}$  is the predicted value of species  $j$  in cell  $i$ , and  $\sum_{k \in S} p_{kj}$  is the sum of values of species  $j$  in cells  $k$  included in the remaining set of cells  $S$  at each step in the cell removal process. The output priority rank map has values from 0 (lowest rank) to 1 (highest).

Our region has 3464042 cells with data. We used default Zonation parameters with a warp factor of 1000 and edge removal on, meaning that in each iteration, 1000 cells on the edges of remaining areas are removed. All species were weighted equally. In total, 57 priority maps were produced, each based on the different SDM variants with all 25 species. We also produced 10 priority maps in Zonation with random cell removal and base SDMs as inputs, to serve as a reference point.

### Measuring differences in priority maps

We used two metrics to measure differences in conservation priorities between solutions produced with the base SDMs and each variant: i) absolute difference in priority ranks (Kujala et al. 2018b), describing the overall change in grid cell rankings; and ii) spatial overlap of both top and bottom priority areas. Absolute differences between two Zonation priority maps were calculated as

**Equation 3.**

$$\frac{\sum_{i=1}^S |\text{pri}_i^{\text{base}} - \text{pri}_i^{\text{var}}|}{0.5 \times n}$$

where  $\text{pri}_i^{\text{base}}$  and  $\text{pri}_i^{\text{var}}$  are the priority rankings of grid cell  $i$  in a solution produced using the baseline SDMs and one of the model variants, respectively. The number of grid cells with rankings ( $n$ ) multiplied by 0.5 denotes the maximum possible difference in ranks that takes place when two solutions are the exact mirror images of each other. Eq. 3 returns values between 0 (solutions are identical) and 1 (solutions are mirror images, named opposite hereafter). Since the chosen baseline might affect relative impacts of other variants, we present alternative (binary or truncated) baseline analyses in Supporting Information.

We evaluated the spatial agreement in areas of highest and lowest priority by calculating the overlap of the top and bottom ranked 20%, 10% and 5% of all cells between each model variant and the base.

## Results

### Comparing SDMs and mapped prioritisations

AUCs are a common way to compare SDMs (Fig. 2). As different variants imposed different numbers of presence and background records, most AUCs are not strictly comparable and trends can only be interpreted in broad terms. All variants displayed reasonable model performance and would pass the goodness of fit test of  $AUC > 0.7$  (Morán-Ordóñez et al. 2017). Models based on the same data showed a slight trend towards decreased discrimination (lower AUC) when model complexity or predictor sets were fixed across species, rather than allowed to vary as in the base model.

Visual (Supporting Information) and quantitative (Fig. 2) comparison of maps revealed larger differences in SDM predictions than suggested by AUCs. Spatially, dealing (or not dealing) with bias had notable impacts on predictions, particularly in areas with less records. Imposing a binary threshold aggregated areas of high predictions. Correlations (COR) broadly matched the two sides of the asymmetric KUL, but with differences in details. Of the SDM variants, complex, bias grid, and truncated were all similar to the base model, measured by COR and either KUL. Binary thresholds had a median COR of 0.75 with base predictions (Fig. 2b), and compared with the base, demonstrated a high agreement of top values (median base-to-var KUL = 0.2). However, from the var-to-base KUL viewpoint (Fig. 2d), binary predictions were the most different to base predictions, presumably because all the nuances about intermediate predictions are missing in the binary version. Not treating bias and varying the number of background points meant some high predictions, especially with more background points, were similar to the base (low var-to-base KUL), but discrepancies were large for many other medium to high predictions (high base-to-var KUL). This

fed through to the typical variant, which was one of the more different variants compared with the base. In summary, many of the SDM variants altered predictions strongly enough to show quantifiable differences compared with the base models.

Differences in SDMs fed through to differences in mapped prioritisations (Fig. 3). There were several similarities between the predictions (Fig. 2, Supporting Information) and prioritisation solutions (Fig. 3), highlighted by the large high priority areas in the central north and north-east, southern tip and extending to north-west. However, there were also marked differences. The grassy north-western tip and more cleared central regions were particularly sensitive to changes in bias treatment, coinciding with lower densities of observations across all species (Fig. 3a,e-h). Additionally, the extent to which the central west (dominated by dry sclerophyll forests, Fig. 1) is valued was highly variable between model variants.

#### **Absolute difference in Zonation ranks**

Absolute differences in rank values (Fig. 4, Supporting Information) were largest when sampling bias was not treated or binary thresholds were applied. Not dealing with bias, regardless of number of background points, resulted in the largest change in rank values from any single SDM variant. Varying the number of background points had a minor effect on the absolute difference, exhibiting a standard deviation of just 0.02% between the three models. Binary thresholding of base SDMs resulted in, on average, 34% of cells being assigned an opposite priority rank value due to this change alone. In comparison, applying a truncated threshold to the base models resulted in a difference of just 5.5%, the lowest of any single model variation. When combined with other variants of model building, truncated thresholds added very little to the deviation from base prioritisation, whereas the application of a binary threshold showed greater additive effects on rankings, particularly when main model effects were less severe (Fig. 4). Additive effects of multiple model changes (typical model) resulted in the largest deviation of all variants. All variants changed the

priority rankings less than an entirely random solution (average rank difference 67%), however, the largest alterations reached approximately two-thirds of the effect of randomization.

### **Overlap in top and bottom priority sites**

Overall, the overlap of top and bottom priority areas between solutions mimicked results seen in rank differences, with least overlap between base and models without bias adjustments or with binary-thresholded predictions (Fig. 5). Focusing on either top or bottom ranked areas revealed more nuanced differences. Not accounting for sampling bias particularly changed bottom priority sites, to the point that the further effects of thresholding were masked, and the overlap approached a random prioritisation (average 12% for bottom priorities). In models accounting for bias, binary thresholding strongly altered the location of all priority areas, while truncated thresholding effects were mild. The mildest effects, resulting from alternative bias treatment and a reduction in species data, preferentially altered the top priority sites, particularly when paired with thresholding. Making MaxEnt models simpler resulted in lower overlap with the base, compared with increased complexity.

Overall, the smaller the priority fraction being examined (e.g. top or bottom 5%), the more sensitive it was to SDM changes. Across model alterations (excluding thresholding), most drastic single changes were seen in bottom ranks, although on average top and bottom ranks were fairly equally affected. For priority maps see Supporting Information.

### **Discussion**

Conservation practitioners are often concerned about data and model uncertainties, and their impact on biodiversity outcomes. But from a decision-making point of view, reducing uncertainty is only relevant if it changes the management decision at hand (Raiffa & Schlaifer 1961).

Our results show that different decisions made during model building are unequal in their impact. Converting continuous predictions into binary maps via thresholds or not accounting for data bias result in the largest alterations in conservation decisions. Decisions about model complexity and predictor variables had a smaller but not insignificant effect, whereas altering the number of background points seemed to carry little effect beyond decisions about bias treatment. Biases are pervasive in the species databases commonly used for SDM building (Merow et al. 2013), and published SDMs are still frequently built without any consideration of bias (Tulloch et al. 2016). Our findings (Fig. 4 and 5) emphasise the large impacts of accounting for bias which, based on previous research (Phillips et al. 2009; Fithian et al. 2015), likely improves prediction accuracy.

Similarly, large changes to spatial plans can be introduced by converting continuous predictions into binary maps, which in our example caused a 34% deviation in cell ranks. Binary thresholding effectively discards information in species distribution predictions (Wilson et al. 2005; Elith & Leathwick 2009b) and has two main effects on the prioritisation: first, information on the relative environmental suitability of grid cells is lost. Cells above the threshold value are all seen as equally valuable and marginal habitats can be prioritised over core ones, as there is no information to distinguish between them. Second, flattening continuous predictions into single value (either 1 or 0) increases ties in the priority ordering of grid cells. These are commonly solved by enforcing random ordering, as is the case with Zonation (Moilanen et al. 2014), which increases randomness in the priority solution. The relationship between suitability and abundance can be non-linear and noisy, but there exists strong evidence that it is a positive relationship and is stronger when models are fitted with covariates beyond simple climatic ones (Weber et al. 2017), as we have done. Previously, Guillera-Aroita et al. (2015) have illustrated the negative effects of thresholding on resource utilisation and target hitting in conservation, arguing against their use. Similarly, Calabrese et al. (2014) show the negative impact of thresholding in stacking SDMs for richness estimates. In applications that require delineation of habitat patches (e.g. population simulations) thresholding

may be justified, but in spatial multi-species prioritisations, thresholding is typically not needed and should not be done without justification. Where necessary, we recommend the use of truncated thresholding, which has minimal effect in comparison to the use of binary maps.

Alterations in SDMs also affect locations of top and bottom priority areas differently. Not accounting for bias altered bottom ranked areas the most, shifting them into less well surveyed areas (Fig. 1 and 3), whereas other SDM variations, such as variable selection, change top priorities more. Hence, alternative model building and fitting options affect decisions differently depending on prioritisation use. Where used to guide placement of new developments to avoid important biodiversity areas (i.e. looking for lower ranked sites), not accounting for bias is likely to introduce largest uncertainty, equivalent to an entirely random solution (Fig. 5).

Our results also indicate that fewer species records do not necessarily introduce large uncertainty in spatial conservation prioritisations. Although the adequacy of species observation data to build reliable model predictions is often seen as a key concern (Elith & Leathwick 2009a; Bean, Stafford & Brashares 2012), halving the number of records only mildly altered spatial priorities (Fig. 4), and retained reasonable (on average 64%) overlap with baseline priority areas (Fig. 5). The tight range amongst iterations suggests that the uncertainty is not strongly affected by the specific species points included, rather the reduction of information overall. Despite halving the data, only four species dipped below 30 records, suggesting there may still have been enough information to build a reasonable model (Wisz et al. 2008). As the default MaxEnt settings for feature classes adjusts model complexity based on number of records, MaxEnt is perhaps particularly effective with handling smaller datasets (Elith et al. 2011), so these findings might not directly apply to SDMs more broadly. Yet, the weak effect of altering species records on spatial priorities has also been shown elsewhere (Grantham et al. 2008; Selwood et al. 2019). As we used cleaned data and accounted for bias based on full datasets, our sub-setting may show reduced effects compared with cases where only few poorer quality records exist. The difference between less records and bias or thresholding is

nevertheless interesting, particularly as practices for handling bias or using thresholds are easier to change than collecting more species records through field surveys.

We emphasise that we did not have independent, unbiased data to test these model variants, and therefore were not aiming to identify the best way to fit SDMs. Rather, we tested alternatives that reflect common practice. There are many approaches for fitting models that could be tested in the context of conservation planning. These include search optimisations across many combinations of feature classes and regularisation parameters in MaxEnt (Cobos et al. 2019) as well as testing across multiple modelling methods or using ensembles (Hao et al. 2020).

Point-based SDM evaluations, exemplified by AUC (Fig. 2a), gave little insight on the extent to which SDM decisions impact prioritisations. These did not indicate substantial differences between variants, and if used to choose a best approach for prioritisation, would have selected no bias treatment with 30000 background points. This is not a problem with the metric (AUC); the issue is that the data are biased, so models that are also biased (i.e. no bias treatment) appear better when tested with the biased data (El-Gabbas & Dormann 2018). Correlations between alternative SDM variants gave more nuanced results but were still limited indicators of decision uncertainty (Fig. 2b). Kulczynski's coefficient successfully flagged some large impacts (Fig. 2c, d), but can be difficult to interpret, since outcomes depended on which way the comparison was made. These results indicate that there is scope for developing additional spatial metrics for comparing mapped predictions.

Spatial prioritisation results are driven by many factors, notably the conservation objective and constraints posed on optimisations. Our results must be interpreted with these in mind. We only considered effects of SDM variations using one spatial planning algorithm. We chose Core Area Zonation because it allows changes in priorities to be most transparently linked to SDM alterations. Additive algorithms (e.g. ABF Zonation) are further influenced by data-specific richness patterns, making it harder to separate impacts of SDM choices from those of data characteristics. Under strong richness gradients, use of additive algorithms can reduce the differences between SDM

effects (Moilanen et al. 2013) but with the cost of reducing protection for species in species-poor habitats. Similarly, target-based algorithms (e.g. Marxan) require defined species' targets, another subjective decision which also constrain the solution. A large caveat is that we do not know the true distributions of species. Some effects may change with a different baseline, although we found our results to be stable, at least against alternative thresholding approaches applied on the baseline (Supporting Information). We therefore believe our assessment about relative magnitudes of differences is informative and a useful starting point for further explorations, including simulations with known truth.

Accounting for data biases and not thresholding predictions unless necessary are good first steps in reducing model-based uncertainty in spatial plans. But cases where multiple equally valid pathways to SDM building exist require more cautionary approaches. For example, in our analysis using an alternative bias correction still produced notable changes in spatial priorities. When no clear reasons to choose one option over another exists, it is important to explore the full range of uncertainty and consider its effects on the decision at hand (Smith et al. 2019). A robust way could be to produce solutions with both options and select sites that are priorities irrespective of the method used (Moilanen et al. 2006). Alternatively, additional work to identify the more accurate choice could be undertaken.

Model predictions are useful, but they will never perfectly represent the true landscape. When using SDMs for decision making, it is important they are the best estimates we can produce, limiting uncertainty as much as possible (Barry et al. 2006). The average 30% difference in prioritisation outcomes resulting from our model alterations is a clear indicator of the impact of decisions made during model production. Consistent with other recent calls for reproducibility (Araújo et al. 2019), our findings demonstrate that model production decisions should not be made arbitrarily, but clearly reported and justified.



## Acknowledgments

This work was supported by the Australian Research Centre Discovery Project grant DP160101003. HK also acknowledges the support of Australian Commonwealth's National Environment Science Program through the Threatened Species Recovery Hub and the Academy of Finland Strategic Research Council project IBC-Carbon, grant #312559. We thank A. Whitehead, who compiled the original species data records and environmental variables, and J.J. Lahoz-Monfort for discussions on the topic. AM extends personal gratitude to those who supported her through the production of this paper.

## Supporting Information

Details of species used in the study (Appendix S1), environmental variables (Appendix S2), soil classes (Appendix S3), correlation between environmental variables (Appendix S4), environmental variables allocated to taxonomic groups (Appendix S5), further details of TGB and bias layer construction (Appendix S6), SDM maps for *Antechinus flavipes* (Appendix S7), further analysis on the effect of thresholding on rank values (Appendix S8) and mapped top and bottom ranked Zonation priority areas (Appendix S9) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

## Literature Cited

- Araújo MB, Anderson RP, Barbosa AM, Beale CM, Dormann CF, Early R, Garcia RA, Guisan A, Maiorano L, Naimi B. 2019. Standards for distribution models in biodiversity assessments. *Science Advances* **5**.
- Atlas of Living Australia. 2020. Available from <http://www.ala.org.au> (accessed May 2018).
- Ball IR, Possingham H, Watts M 2009. *Marxan and relatives: Software for spatial conservation prioritisation*. Oxford University Press, Oxford, UK.
- Barry S, Elith J. 2006. Error and uncertainty in habitat models. *Journal of Applied Ecology* **43**:413-423.
- Bean WT, Stafford R, Brashares JS. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography* **35**:250-258.
- Calabrese, J.M., Certain, G., Kraan, C., Dormann, C.F., 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography* **1**:99-112.
- Cobos, M.E, Townsend Peterson, A, Osorio-Olvera, L, Jiménez-García, D. 2019. An exhaustive analysis of heuristic methods for variable selection in ecological niche modeling and species distribution modeling. *Ecological Informatics* **53**.
- Colin MB, Jack JL. 2012. Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions: Biological Sciences* **367**:247-258.
- El-Gabbas A, Dormann CF. 2018. Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. *Ecography* **41**:1161-1172.
- Elith J. 2019. Chapter 15: Machine-Learning, Random Forests and Boosted Regression Trees, editor Brennan L.A., Tri, A. N. & Marcot, B. G. *Quantitative Analyses In Wildlife Science*. John Hopkins University Press.

- Elith J, Leathwick JR. 2009a. The Contribution of Species Distribution Modelling to Conservation Prioritization. Pages 70-93 in Moilanen A, Wilson KA, and Possingham HP, editors. *Spatial Conservation Prioritization: Quantitative Methods & Computational Tools*. Oxford University Press, Oxford, UK.
- Elith J, Leathwick JR. 2009b. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics* **40**:677-697.
- Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and distributions* **17**:43-57.
- Fithian W, Elith J, Hastie T, Keith D. 2015. Bias correction in Species Distribution Models: Pooling Survey and Collection Data for Multiple Species. *Methods in Ecology and Evolution* **6**:424-238.
- Foody GM. 2011. Impacts of imperfect reference data on the apparent accuracy of species presence–absence models and their predictions. *Global Ecology and Biogeography* **20**:498-508.
- Grantham HS, Moilanen A, Wilson KA, Pressey RL, Rebelo TG, Possingham HP. 2008. Diminishing return on investment for biodiversity data in conservation planning. *Conservation Letters* **1**:190-198.
- Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* **24**:276-292.
- Guisan A, Thuiller W, Zimmermann NE 2017. *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press.
- Guisan A, et al. 2013. Predicting species distributions for conservation decisions. *Ecology Letters* **16**:1424-1435.

- Hao T, Elith J, Lahoz-Monfort J.J, Guillera-Arroita G. 2020. Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography* **43**:549-558.
- Hermoso V, Kennard MJ. 2012. Uncertainty in coarse conservation assessments hinders the efficient achievement of conservation goals. *Biological Conservation* **147**:52-59.
- Kujala H, Moilanen A, Gordon A. 2018a. Spatial characteristics of species distributions as drivers in conservation prioritization. *Methods in Ecology and Evolution* **9**:1121-1132.
- Kujala H, Lahoz-Monfort JJ, Elith J, Moilanen A. 2018b. Not all data are equal: Influence of data type and amount in spatial conservation prioritisation. *Methods in Ecology and Evolution* **9**:2249–2261.
- Kujala H, Whitehead AL & Wintle BA. 2015. Identifying conservation priorities and assessing impacts and trade-offs of potential future development in the Lower Hunter Valley in New South Wales. NERP Environmental Decisions Hub report. The University of Melbourne, Melbourne, Victoria.
- Langford WT, Gordon A, Bastin L. 2009. When do conservation planning methods deliver? Quantifying the consequences of uncertainty. *Ecological Informatics* **4**:123-135.
- Legendre L, Legendre P 1998. *Numerical Ecology. Developments in Environmental Modelling*. 2<sup>nd</sup> Edition. Elsevier, New York.
- Merow C, Smith MJ, Silander JA. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* **36**:1058-1069.
- Merow C, Smith MJ, Edwards TC, Guisan A, McMahon SM, Normand S, Thuiller W, Wüest RO, Zimmermann NE, Elith J. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography* **37**:1267-1281.
- Moilanen, A, Anderson, BJ, Arponen, A, Pouzols, FM, & Thomas, CD. 2013. Edge artefacts and lost performance in national versus continental conservation priority areas. *Diversity and Distributions*, **19**:171–183.

- Moilanen A, Franco AM, Early R, Fox R, Wintle BA, Thomas CD. 2005. Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proceedings of the Royal Society B: Biological Sciences* **272**:1885-1891.
- Moilanen A, Pouzols F, Meller L, Veach V, Arponen A, Leppänen, Kujala H. 2014. Spatial conservation planning methods and software Zonation Version 4 User manual. C-BIG Conservation Biology Informatics Group, Helsinki, Finland.
- Moilanen A, Runge MC, Elith J, Tyre D, Carmel Y, Fegraus E, Wintle BA, Burgman MA, Ben-Haim Y. 2006. Planning for robust reserve selection using uncertainty analysis. *Ecological Modelling* **199**:115-124.
- Morán-Ordóñez A, Lahoz-Monfort JJ, Elith J, Wintle BA. 2017. Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions?. *Global Ecology and Biogeography* **26**:371-384.
- NSW Office of Environment and Heritage. 2020. BioNet Atlas. Available from <http://www.bionet.nsw.gov.au/> (accessed May 2018).
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**:231-259.
- Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**:161-175.
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**:181-197.
- Raiffa H, Schlaifer R. 1961. Applied statistical decision theory. Harvard University, Boston.
- Ray N, Burgman MA. 2006. Subjective uncertainties in habitat suitability maps. *Ecological Modelling* **195**:175-186.
- Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI. 2015. Point process models for presence-only analysis. *Methods in Ecology and Evolution* **6**:366-379.

- Selwood KE, Wintle BA, Kujala H. Collaborative conservation planning: Quantifying the contribution of expert engagement to identify spatial conservation priorities. *Conservation Letters* **12**.
- Smith AL, et al. 2019. Managing uncertainty in movement knowledge for environmental decisions. *Conservation Letters* **12**.
- Synes NW, Osborne PE. 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Global Ecology and Biogeography* **20**:904-914.
- Tulloch AI, Sutcliffe P, Naujokaitis-Lewis I, Tingley R, Brotons L, Ferraz KMP, Possingham H, Guisan A, Rhodes JR. 2016. Conservation planners tend to ignore improved accuracy of modelled species distributions to focus on multiple threats and ecological processes. *Biological Conservation* **199**:157-171.
- Weber, M.M., Stevens, R.D., Diniz-Filho, J.A.F., Grelle, C.E.V. 2017. Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. *Ecography* **40**:817–828.
- Williams KJ, Belbin L, Austin MP, Stein JL, Ferrier S. 2012. Which environmental variables should I use in my biodiversity model? *International Journal of Geographical Information Science* **26**:2009-2047.
- Wilson KA, McBride MF, Bode M, Possingham HP. 2006. Prioritizing global conservation efforts. *Nature* **440**:337-340.
- Wilson KA, Westphal MI, Possingham HP, Elith J. 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation* **122**:99-112.
- Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham C, Guisan A, Group NPSDW. 2008. Effects of sample size on the performance of species distribution models. *Diversity and distributions* **14**:763-773.

**Table 1.** Data, settings and approaches used in the alternative SDMs produced in this study.

Model variant name	feature types	number of background points	approach to bias	environmental variable selection (number of variables) (Supporting Information)	species data used
Base	α) autofeature on	β) birds- 17833 mammals- 20730 plants- 11904 reptiles- 5341	γ) target group background samples (Supporting Information)	δ) vary with taxonomic group (16 to 19)	ε) all
Complex	enforce full complexity (autofeature off, threshold on)	β	γ	δ	ε
Simple	Linear and quadratic only	β	γ	δ	ε
Predictors	α	β	γ	same set for all species (12)	ε
Bias grid	α	20000	selection of background points weighted by bias layers	δ	ε
No bias 2	α	2000	none- random background points used	δ	ε
No bias 10	α	10000	none- random background points used	δ	ε
No bias 30	α	30000	none- random background points used	δ	ε
Few records (10 iterations)	α	β	γ	δ	random 50% sample, 10 repeats
Typical	α	10000	none- random background points	same set for all species (12)	ε
Binary	a threshold was used to convert predictions to make binary or truncated predictions; applied to all SDM variants				
Truncated					

## Figure legends

**Figure 1.** (a) Observation points for 25 species across the Greater Hunter study region in NSW, Australia. Major regional city Newcastle, and the altitudinal peak Brumlow Top within Barrington Tops National Park are shown. (b) Major vegetation types. The x and y axes indicate easting and northing (metres) in the Australian Albers coordinate reference system. The lower inset demonstrates the study site's location within Australia.

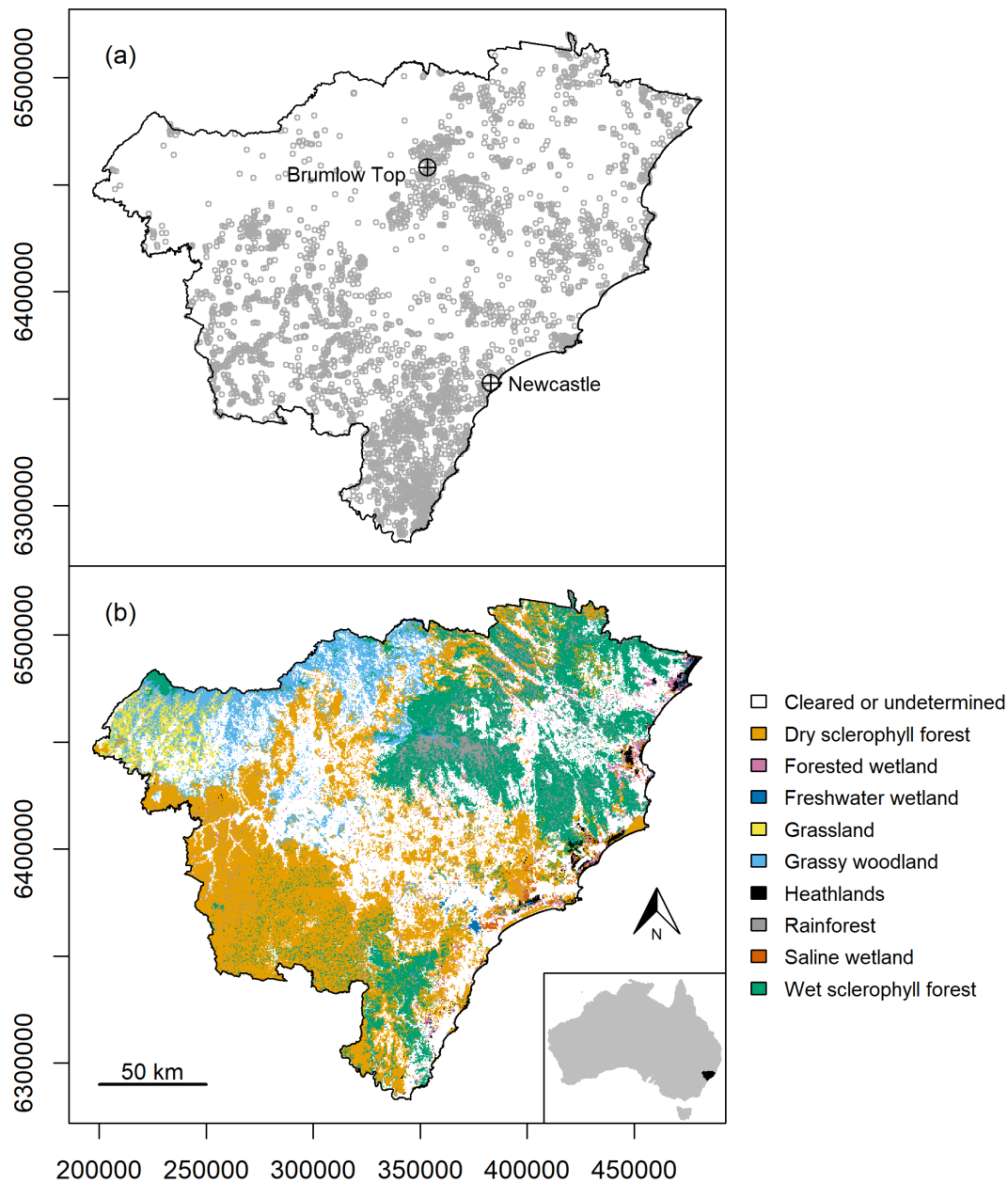
**Figure 2.** Quantitative measures of differences between species distribution model variants for all species: a) Test area under the curve (AUC) calculated across 7-fold cross validation, b) Pearson's pairwise correlation coefficient, and Kulczynski's asymmetric coefficient comparing c) base to other variants and d) variants to the base. Box represents the interquartile range with horizontal black line and error bars correspond to median and standard deviation, respectively. Circles represent outliers.

**Figure 3.** Conservation prioritisations produced from the SDMs for all species under different model variants. Colour gradient represents priority rank of each cell across the landscape measured between 0 (least important) and 1 (most important) in terms of biodiversity conservation value.

**Figure 4.** Absolute difference in cell rank values between solutions produced with the base SDMs and each model variant, ordered based on increasing rank difference when no thresholding is used. The bars give the difference between priority solutions when SDMs have not been post-processed (no thresholding, dark grey), and when they have been modified with a truncated (intermediate grey) or binary (light grey) thresholding approach. Error bars represent standard deviation from the mean for solutions with multiple iterations.

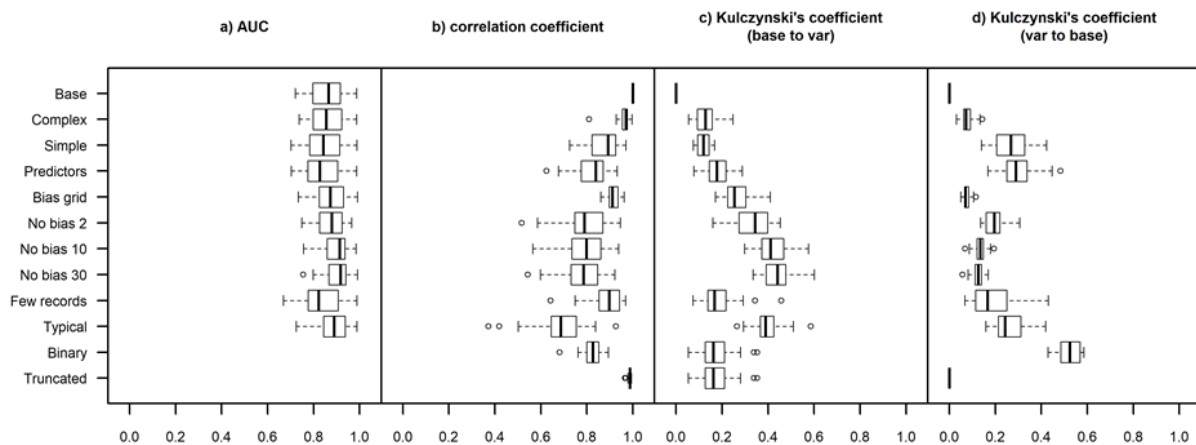
**Figure 5.** Overlap of the top and bottom 20%, 10% and 5% priority sites, comparing prioritisations based on each species distribution model variant to that with the base model. Comparisons are segregated by versions of model variants with (a) continuous prediction with no post-processing, (b) truncated thresholding and (c) binary thresholding. Error bars represent standard deviation from the mean for solutions with multiple iterations.



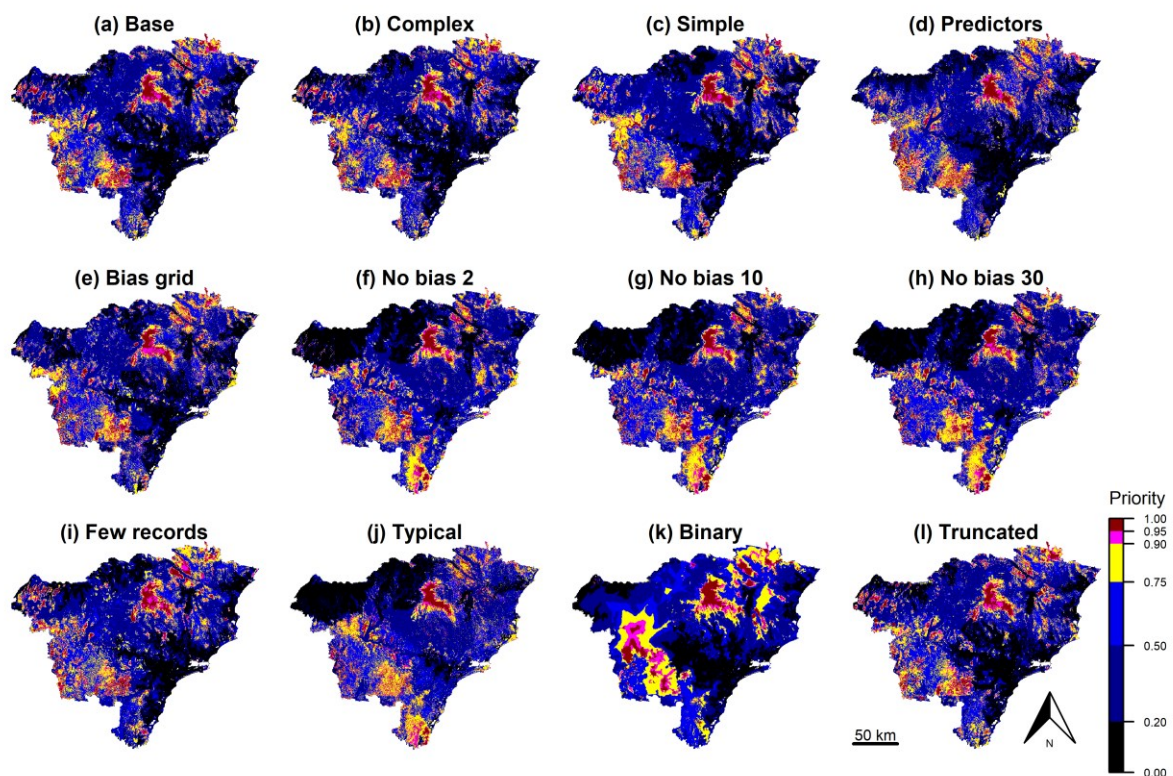


**Figures**

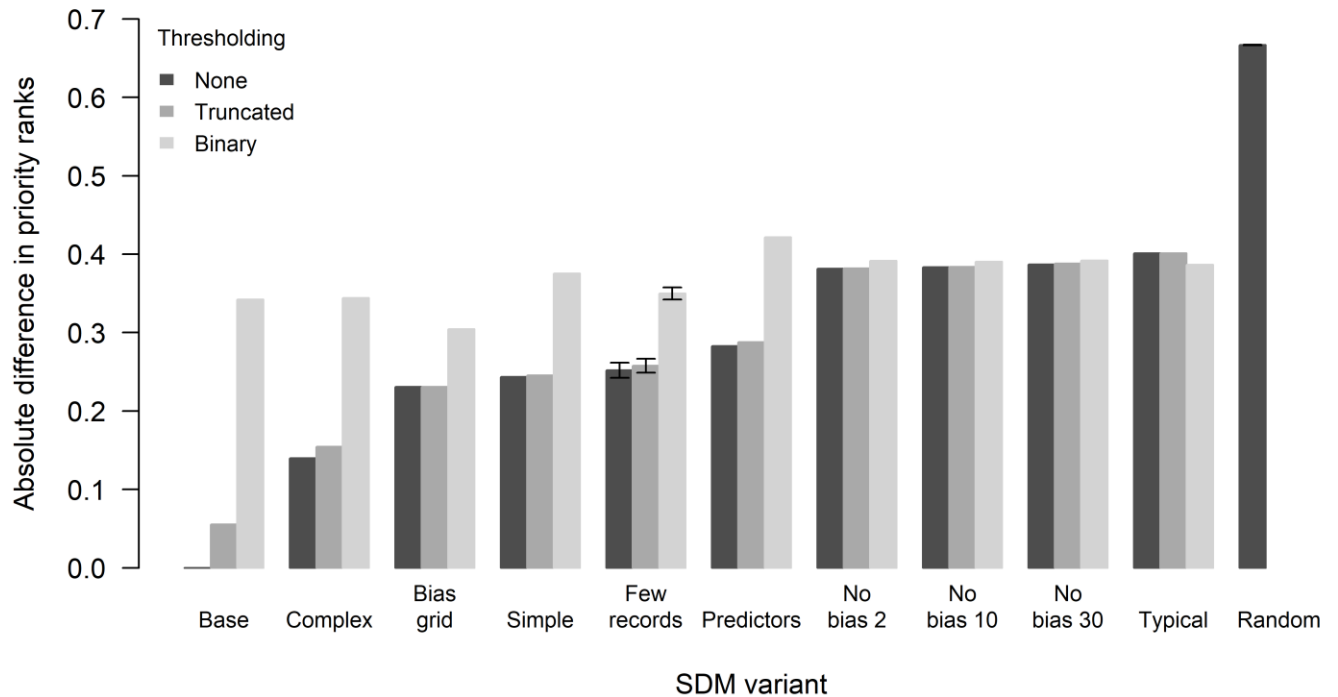
**Figure 1.** (a) Observation points for 25 species across the Greater Hunter study region in NSW, Australia. Major regional city Newcastle, and the altitudinal peak Brumlow Top within Barrington Tops National Park are shown. (b) Major vegetation types. The x and y axes indicate easting and northing (metres) in the Australian Albers coordinate reference system. The lower inset demonstrates the study site's location within Australia.



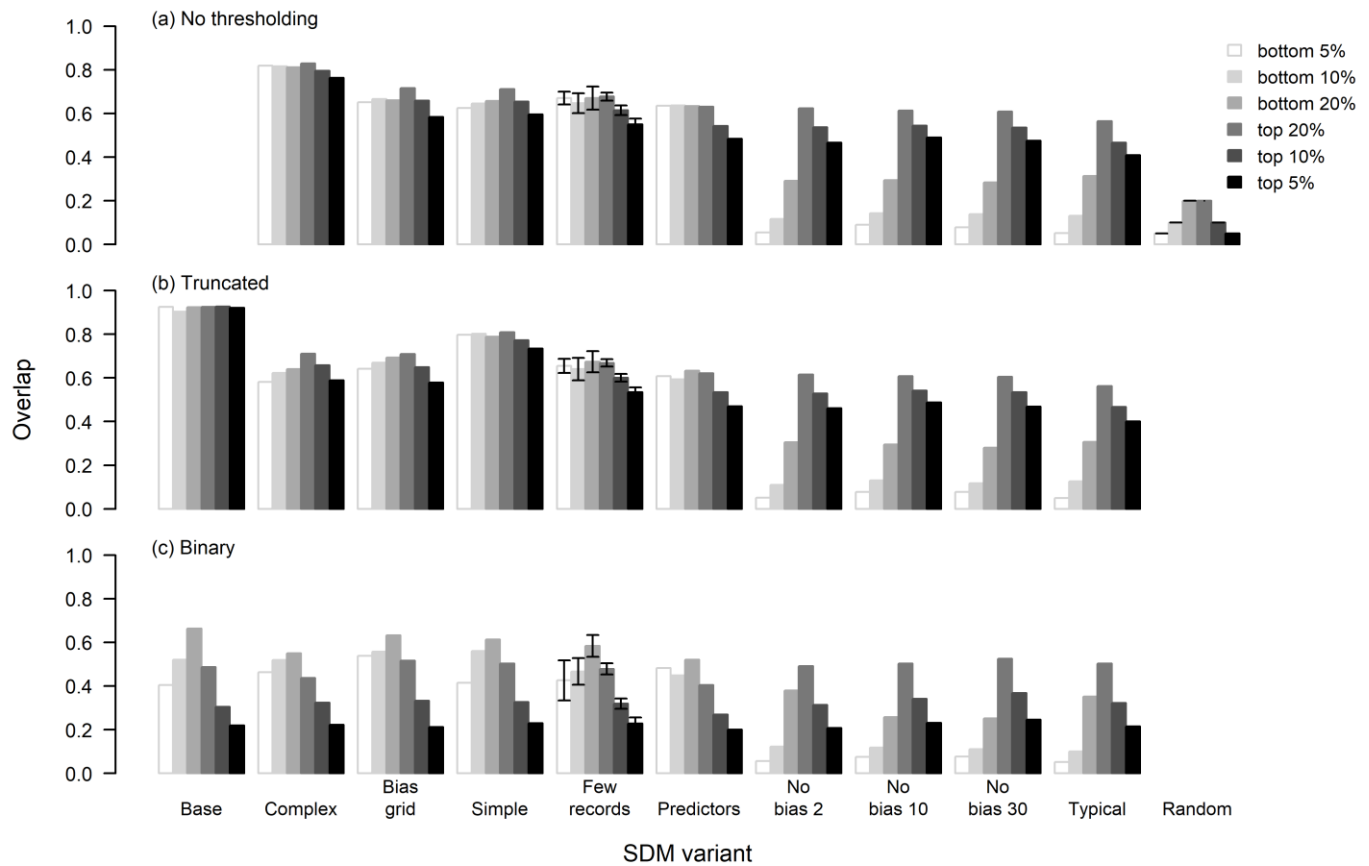
**Figure 2.** Quantitative measures of differences between species distribution model variants for all species: a) Test area under the curve (AUC) calculated across 7-fold cross validation, b) Pearson's pairwise correlation coefficient, and Kulczynski's asymmetric coefficient comparing c) base to other variants and d) variants to the base. Box represents the interquartile range with horizontal black line and error bars correspond to median and standard deviation, respectively. Circles represent outliers.



**Figure 3.** Conservation prioritisations produced from the SDMs for all species under different model variants. Colour gradient represents priority rank of each cell across the landscape measured between 0 (least important) and 1 (most important) in terms of biodiversity conservation value.



**Figure 4.** Absolute difference in cell rank values between solutions produced with the base SDMs and each model variant, ordered based on increasing rank difference when no thresholding is used. The bars give the difference between priority solutions when SDMs have not been post-processed (no thresholding, dark grey), and when they have been modified with a truncated (intermediate grey) or binary (light grey) thresholding approach. Error bars represent standard deviation from the mean for solutions with multiple iterations.



**Figure 5.** Overlap of the top and bottom 20%, 10% and 5% priority sites, comparing prioritisations based on each species distribution model variant to that with the base model. Comparisons are segregated by versions of model variants with (a) continuous prediction with no post-processing, (b) truncated thresholding and (c) binary thresholding. Error bars represent standard deviation from the mean for solutions with multiple iterations.